DOCUMENT RESUME

ED 221 577 TM 820 606

AUTHOR Fleming, Margaret; And Others
TITLE Standardized Testing under Court Ordered Desegregation: Developing a Compliance Plan.
INSTITUTION Cleveland Public Schools, Ohio.
PUB DATE Mar 82
NOTE 75p.; Papers presented at the Annual Meeting of the National Council on Measurement in Education (New York, NY, March 20-22, 1982).

EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS Advisory Committees; *Compliance (Legal); *Court Litigation; Criterion Referenced Tests; Culture Fair Tests; *Desegregation Plans; Elementary Secondary Education; Equal Education; Inservice Teacher Education; *School Districts; Standardized Tests; Teacher Made Tests; *Test Bias; Test Construction; Testing Programs; *Test Use
IDENTIFIERS *Cleveland Public Schools OH

ABSTRACT

Court ordered desegregation introduces significant requirements for school districts' testing programs. The Remedial Order in effect in Cleveland, Ohio (Reed vs. Rhodes) cautions defendants that "they must insure that all tests, whether standardized, criterion-referenced, or teacher-made are developed, administered, and scored in a nondiscriminatory manner." The court mandate includes teacher-made tests as well as standardized tests. A compliance plan must ensure that standardized and classroom tests be used in a nondiscriminatory manner. Bias in instruments and testing procedures must also be monitored at all levels. The Cleveland School District developed a compliance plan which includes a Test Review Committee to select and monitor the testing program, staff development programs using task-relevant materials and peer teams to demonstrate diagnostic use of standardized tests, validated slide-tape packages and hands-on materials to train teachers to improve classroom tests, and student diagnostic profiles together with individualized narratives to facilitate dissemination of results to students and parents. For each component, a description of service, delivery procedure, and assessment findings is discussed. First year assessment supports the plan's feasibility and effectiveness in fulfilling the letter and spirit of the desegregation order. (Author/CM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED221577

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

STANDARDIZED TESTING UNDER COURT ORDERED DESEGREGATION: DEVELOPING A COMPLIANCE PLAN

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

M. Fleming

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Margaret Fleming, Chairman
Marian Karantanes
Jacqueline Reed
Janie Rollins
Hazel Redden
Barbara Chambers

- Cleveland Public Schools -

TM 820 606

Symposium presented to annual meeting of the National Council on Measurement in Education, New York, N.Y., March, 1982.

- ABSTRACT -

Title: STANDARDIZED TESTING UNDER COURT ORDERED DESEGREGATION: DEVELOPING A COMPLIANCE PLAN

Court ordered desegregation introduces significant requirements for school districts' testing programs. The Remedial Order in effect in Cleveland, Ohio (Reed vs. Rhodes) cautions defendants that "they must insure that all tests, whether standardized, criterion-referenced, or teacher-made are developed, administered, and scored in a nondiscriminatory manner." The court mandate includes teacher-made tests as well as standardized tests. A compliance plan must ensure that standardized and classroom tests be used in a nondiscriminatory manner. Bias in instruments and testing procedures must also be monitored at all levels.

The Cleveland School District developed a compliance plan which includes a Test Review Committee to select and monitor the testing program, staff development programs using task-relevant materials and peer teams to demonstrate diagnostic use of standardized tests, validated slide-tape packages and hand-on materials to train teachers to improve classroom tests and student diagnostic profiles together with individualized narratives to facilitate dissemination of results to students and parents. For each component, a description of service, delivery procedure, and assessment findings will be discussed. First year assessment supports the plan's feasibility and effectiveness in fulfilling the letter and spirit of the desegregation order.

STANDARDIZED TESTING UNDER COURT ORDERED DESEGREGATION: DEVELOPING A COMPLIANCE PLAN

Margaret Fleming
Cleveland Public Schools

Paper presented to annual meeting of the National Council on Measurement in Education, New York, N.Y., on March 21, 1982.

INTRODUCTION

During the past two years, the Cleveland City School District has been implementing a desegregation plan as a result of liability findings in Reed vs. Rhodes. The Remedial Order which addresses remedying the constitutional violations in the segregated Cleveland School District mandates that the district implement specific educational programs. These programs, as ancillary relief, seek to correct the effects of prior segregated schooling and to assure an effective transition to a racially neutral, nondiscriminatory, unitary school system. The Court in its remedial order proposed the provisions with these major purposes:

1. To remedy the academic effects of prior segregation;
2. To ensure that existing and future programs are administered in a nondiscriminatory fashion;
3. To maintain a secure, integrated school environment in which the rights of all students are protected; and
4. To guarantee that Court-ordered educational provisions are successfully implemented.[1]

Through the "Testing and Tracking" component of the Court Order, the Court introduces significant requirements for the district's standardized testing program. The Order states:

> ...The record indicates that achievement and ability testing has long been used in the Cleveland School System as an indicator of student progress. While there has been no evidence of discriminatory use of tests in the past, the evidence adduced in this case indicates that testing can become a tool for resegregation. Therefore, the defendants are cautioned that they must insure that all tests, whether standardized, criterion referenced, or teacher made, are developed, administered, and scored in a nondiscriminatory manner.[2]

As can be seen, the Court Order provides impetus toward re-examination of all aspects of the district's procedures in standardized and teacher-made testing. It stipulates that the district achieve compliance in test development, administration, scoring and use.

This presentation describes the processes undertaken to develop a compliance plan to meet the requirements of the Court's Remedial Order. Subsequent presentations in this round-table discussion explain the various components of the compliance effort with a description of service, delivery procedures and assessment findings.

RE-APPRAISAL OF TESTING

A re-appraisal of beliefs about testing, particularly standardized testing and its use as a potential instrument of discrimination is necessary to approach the task. This re-appraisal is guided by key questions:

- Do standardized tests have a place in a desegregating public school system?
- What are major requirements for a non-discriminatory testing program?

A re-appraisal of beliefs about testing, particularly standardized testing, and its use as a potential instrument of discrimination is necessary to approach the task. Whether standardized tests have a place in public schools, particularly those which have been desegregated, related to one's view of education. If education is seen as developing cognitive behaviors with evaluation as part of the instructional fabric in the mode of Ralph Tyler, then the use of tests is self-evident. If education is seen as primarily fostering personal and social behaviors, then tests are considered to be irrelevant and even damaging. Whatever the model of instruction, teachers have had to make judgments or evaluations about the progress

and needs of their students. To do this requires a flexible range of judgmental tools. Be it classroom test, quiz, intuitive hunch or standardized test, the need for standards of validity and reliability in these tools are critical. The standardized test, as these other procedures, if valid and reliable, has potential for providing one element of such judgments.

Although the relationship between valid and reliable testing and effective teaching may be accepted by minority communities, all too frequently actual experience with standardized tests in schools may have demonstrated discriminatory consequences:

- They have learned that the scores may have set limits for their children rather than diagnosing their needs.
- Standardized test scores are seen as molding teachers' expectations for their students and as guiding placement of students in "tracks" which effectively determine the quality of schooling students will receive.

School districts in desegregation require intensive and extensive examination of their testing programs and the educational system which such programs help to maintain.

What are Major Requirements for a Non-Discriminatory Testing Program?

Continuation of tests in the Cleveland School District necessitated close scrutiny of all aspects of testing. As the Court Order established, development, administration, scoring and use were key areas for examination. Efforts of the district in these areas were re-appraised.

Over the years, the district had attempted to maintain credibility of its testing program. Credibility was seen as beginning with "how" a program was selected and installed in the district. Credibility began with test selection. The use of a Test Review Committee to facilitate communication between those that Dyer (1973) identifies

as "the four groups of people who are involved in the transactions which we call educational testing: the test makers, the test givers, the test takers and the test users."[3] As Dyer emphasizes there is a wide range of understanding and opinion among these groups. Communication among the groups, in his opinion, has become increasingly strained and tenuous in the matter of tests. The involvement of these groups through membership and input to the Committee attempted to provide for clear communication and to build a base for consensus.

The Committee studied the development of the tests meeting with publishers' representatiives to examine test rationales and development procedures aimed at elimination of bias in their tests. Coupled with scrutiny of the tests' development was try-out in local classrooms where administration procedures of proposed tests were demonstrated. Try-out provided opportunity for school-based staff--principals and teachers--to identify potential trouble areas in content, format, length, administration, answer sheets and scoring. Additionally, try-out permitted comparison of levels of performance obtained from new and old tests.

Additional facets of credibility for the testing program centered on maintenance of quality control of the testing program. Polemeni (1975) describes viable operation of a testing program as including systematic delivery of tests and related materials, carefully administered testing schedules and vigilant, thorough supervision of test security.[4] Accuracy of the scoring process has been maintained over the years through careful screening of answer sheets submitted from the schools and maintenance of computerized programs of error checks in the scoring process. Timeliness of results has been addressed through rapid turn-around procedures through "no little" effort on the

part of the staff. Given the nature of human affairs in school districts, no effort is spared for careful administration of testing operations. To support administration of tests, bulletins of instructions and test materials such as Directions for Administration, Teacher's Guides and the like have been supplied for every test to all classroom teachers or guidance counselors involved. Recognizing that non-discriminatory administration of tests means provision of equal opportunity for all students in the testing situation, the district has backed up the extensive array of materials for test administration with on-going staff training sessions.

Use of test results presented the most difficult challenge in elimination of discriminatory practice. Over the years the school district has recognized that it would be necessary to improve its output of test results. A continuing concern has been the paper flow to principals, teachers and supervisors which could be intimidating even to the initialed in test interpretation. Careful attention had been given to design of print-outs to facilitate understanding. Few teachers arrive at the classroom with competencies in educational measurement. Actual practice in the matter of print-outs has been a continuing necessity. As Mayo (1967) and others have documented teachers' preservice training largely neglects educational measurement.[5] Yet their needs for such competencies have remainded critical. It is evident that the typical measurement course for teachers according to a recent account by Bradley (1978) is not viewed as relevant by the students, who frequently take them to fulfill a requirement.[6] Preservice education of teachers continues to be ambiguous in its commitment to the utility of evaluation in the instructional cycle and the potential of educational measurement as a facet of that evaluation

process. Inservice training in school districts has followed suit in being just as inconsistent in matters related to educational measurement.

The school district has recognized an obligation, if standardized tests were to be used, to improve capability of the school district staff in dealing with the tests. Unfortunately, continuing controversy about standardized tests has obscured the fact that teachers need to improve their skills in educational measurement. Although professional education associations, community organizations and activist groups demand elimination of standardized tests in the schools, they have given little heed to the improvement of classroom assessment procedures. Ironically, it is the classroom test or quiz or intuitive hunch, for that matter, which provides the teacher with daily, ongoing information about student progress, not the standardized test. Student failure is communicated more frequently by classroom tests than by standardized tests. The fact that the message might come as a result of "faulty" instruments does not seem to concern the critics.

In the matter of use of test results, the most complex of court-clarified requirements, the district has sought to communicate that results for individual students represent, at the very best, a limited sample of students' performance. It has emphasized that tests do not sample all the ways in which students can display knowledge in any subject area; that the test, as a human product, is prone to error and does err; and the use of results has been predicated on tests' potential to provide a measure with appropriate standards of reliability and validity. Another "caveat" which has been amply featured in analysis of test results has been the necessity to incorporate test results into the information base about students, combining such data

with other information. Test data have been considered as supportive information rather than the major factor in decisions about students. The district has been sensitive to the concern, which Levine (1975) and others have identified, that reliance on standardized tests has developed because of the excellence of the tests or because the "tests have served a set of latent social functions which assure...acceptance and limit criticisms?"

In seeking appropriate standards of validity for the city-wide testing program, the district has selected tests on the basis of relationship to the curriculum in use. School people, however, have learned over the years that although there may be a curriculum, a set of guides and even sets of lesson plans available, given conditions in some schools, no uniform curriculum really exists. Critics of schools would consider this to be a blessing! The real condition of curriculum is that it is interpreted by teachers and translated into action by teachers. It may exist in as many versions as there are teachers. Concurrently, managing delivery of instruction to students, if they come at all, in understaffed and poorly financed schools is a challenge. The Court Order in mandating requirements for educational improvement within the desegregation plan provides a powerful impetus for doing things differently to improve educational equity and opportunity.

SEARCH FOR MODELS

Concurrent with the re-appraisal of what had been--came the task of identifying what models were available to achieve compliance with the Court Order. A literature search identified both desegregation and testing literature which appeared to have some relevance for this effort.

In the 28 years since the Supreme Court decision, Brown vs. Topeka Board of Education, an extensive array of desegregation literature has appeared. An overview of this literature shows a progression from social science studies (Clark 1955)[7] and (Coleman et al 1966)[8], which established desegregation policy, to field studies of desegregation, which assessed impact of various mandatory/voluntary plans. The field studies precipitated critical reviews of the research and the findings. Such an example is found in Bradley and Bradley (1977) in their review of 29 field studies concluded since 1959, which concludes that the field studies and the critical reviews provide both support and disconfirming evidence for the use of school desegregation as an intervention strategy.[9] They are of little help in delineating appropriate testing models for use in desegregated settings. Currently, the emphasis of desegregation literature is toward proposals for improvement of desegregation efforts and accompanying evaluation studies. One of these efforts which assesses current knowledge about the effectiveness of school desegregation strategies, the work of Hawley et al (1981), synthesizes findings of some 1,200 studies and experiences of leaders in school districts that have undergone desegregation. Findings from this massive effort of nine volumes show that reassignment of pupils must be buttressed by instructional and organizational changes.

Four of the report's recommendations offer direction for development of a non-discriminatory testing model. These proposals include elimination of grouping of students in separate classes by ability in elementary school; examination of any within classroom ability groups that do not change; elimination of rigid and inflexible tracking and grouping in secondary schools and development of explicit policies and procedures for identifying and placing students in special curriculum in non-discriminatory ways. The report presents alternative assessment procedures that are intended to reduce reliance on standardized intelligence tests. These include:

1. criterion-referenced assessment to describe a child's test performance in terms of level of mastery of specific skills, rather than in comparison with a normative distribution;

2. learning potential assessment which uses a test-teach-retest paradigm in order to assess the child's actual learning ability and strategy;

3. interdisciplinary assessment to combine the perspectives of a variety of professionals who have worked with the child, including the child's classroom teacher;

4. consultation model prescribed by Johnson (1976) in which the necessity of testing is determined after work has been undertaken in the regular classroom with the child before special education placement testing is done; and

5. systems of multicultural pluralistic assessment such as that prescribed by Mercer and Lewis (1978) which attempts to control for different sources of bias through pluralistic norms, ecological assessment of adaptive behavior and interdisciplinary process with emphasis on parent involvement.[10]

Mercer's efforts to develop non-discriminatory assessment is very relevant to this effort. Findings from her dozen years or so of research on how community agencies, particularly the public schools, identified and labeled persons as mentally retarded led her

to conclude that the higher failure rate of children from lower socio-economic and ethnic minority backgrounds on so-called measures of intelligence was a significant factor in producing higher rates for mental retardation in these groups. Finding that adaptive behavior (a child's ability to cope in a variety of social systems), the IQ test and physical disability scores were relatively independent dimensions, Mercer and Lewis (1978) produced a system of multicultural pluralistic assessment (SOMPA).[11] SOMPA utilizes concepts from three assessment models--the medical model, the social system and the pluralistic model, also called the "general intelligence" model. According to Mercer, SOMPA looks at the child's performance from the perspective of each of the conceptual models. A wide variety of patterns of scores is possible and can be used to gain an assessment of the child that is not racially or culturally discriminatory. The SOMPA is most useful in laying the broad diagnostic foundation on which specific interventions can be built.

The testing literature abounds in articles addressing standards for testing practice and bias in measurement instruments. Selected references offer additional guidelines for this mandate from the Court.

A renewal effort in school assessment, though not within the framework of desegregation, proposed by the Management Standards Study Advisory Committee of the Massachusetts Advisory Council on Education (1975) features five assessment methods. The Committee proposed five assessment methods which include not only objective and essay examinations but use of testimony, committee review, performance assessment and product assessment. The general management principles promoted by the Committee appear to be conducive to non-discriminating approaches.

1. A high school or any school community should define clearly the learning goals....State goals should serve as a general starting point for local efforts in this regard.

2. To the degree possible, the school should establish clear and realistic standards for individual performance in relation to both common and individualized learning goals.

3. The school community should match methods of assessing individual performance to the learning environment and to the type of learning environment under consideration.

4. The school should use assessment information both to guide students toward higher levels of achievement and to guide changes in program or learning options....[12]

Criticisms of standardized tests should be weighed as school district testing activities are re-appraised. Houts (1977) compiled 29 articles attacking intelligence and achievement tests which describe the misuse of intelligence and achievement tests.[13] Alternatives to current testing practices are presented. Predictably, the alternatives exhibit the same weaknesses as the intelligence and achievement tests which they are meant to replace.

Since the National Conference on Achievement Testing and Basic Skills in March 1978, the National Institute of Education has conducted four types of activities related to test use which offer potential guidelines for improvement of testing programs: providing information and referral services about achievement tests; conducting regional conferences similar to the national conferences; training teachers, administrators and parents in the use of achievement tests and conducting research to improve testing. As evident in such products of this NIE effort such as the trainer's manual, Training Citizen Groups in Educational Testing (Elford et al 1980) and Your Child and Testing (Herndon 1980), certain key issues must be addressed by school districts in implementing standardized testing programs. These include fairness,

validity, reliability and the legal rights for parents and students in the testing process. Parents have a right to inspect records, a right to privacy, a right to parent involvement in decisions about the use of tests to plan special education programs and a right to fairness in testing. Although these provisions apply particularly to placement in special education programs and the tests utilized in making such decisions, guidelines for standardized testing programs used to assess student progress and to determine impact of the curriculum should also comply with such provisions.

The Task Force on the Code for Tests and Testing (NAACP 1976) calls for the testing industry and the user, not only to engage in introspection, but to adopt a code to work with tests and test-related activities.[16] The recommended code focuses on construction of tests, test standardization, test administration, limitations of test uses, clear interpretation of results, constructive uses of tests, research and improved training of users in test interpretation.

Another area of testing literature which appears pertinent to re-appraisal of standardized testing practices are studies dealing with racial and cultural bias. Some selected references which illustrate issues and concerns in this area include Flaugher (1974)[17] and Merz (1980).[18] Flaugher reviews bias in testing, identifying three categores--bias in test content, bias in administration and bias in use. Flaugher concludes that descriptions of validation statistics showing that minorities are predicted as well as majorities coupled with the essentially negative results from studies of biased content are no response at all to minority criticisms of testing, no matter what the final results may be. In his view, no other statistical evidence is

needed when minorities are underrepresented in meaningful, lucrative and prestigious positions in the society.

Merz describes and applies several methods of assessing test item bias and examines the concept of fair use of tests. Although the bulk of the paper addresses construction of tests, the practitioner operating a school district program can apply such illustrations to improving judgment of the safeguards reported by test publishers in their development of more valid instruments.

Even more formidable problems for school district testing programs can be gleaned by reference to the burgeoning body of literature in racial and cultural bias in measurement instruments. A recent search of documents under this rubric through Educational Information Resources Center (ERIC) reveals 68 articles encompassing state of the art reviews, test construction models, accountability issues, racism in testing and application of results with various instruments. As these examples demonstrate, recent literature shows that dramatic developments have taken place in addressing bias in tests. A most significant collection of such reports and working papers were presented at the state-of-the-art conference on bias in tests (Berk 1980).[19]

Another area of testing literature that needs to be integrated into achieving compliance for the standardized testing program under the Court Remedial Order deals with appropriate uses of standardized testing. Airasian (1979) proposes that test uses within the classroom include providing diagnostic information about a pupil's or a class's learning problems, assessing learning progress of pupils and classes, providing information about pupils which can help teachers select feasible instructional approaches for pupils and classes, providing data helpful

in guiding pupils and their parents regarding various educational options, predicting future performance, evaluating overall effectiveness of the classroom programs and facilitating decisions involving grouping or classifying students, including "special needs" students. This viewpoint that tests provide additional information about pupils which teachers may use to corroborate their own judgments and perceptions about pupils mirrors the district's testing philosophy.

Airasian identifies issues which are significant for development of guidelines for use of standardized tests in all school districts:

> 1. A test, any test, is a sample or behavior and hence, the scores or ratings derived from tests are subject to error.
>
> 2. A large part of the debate over test use and misuse is best summarized as a situation in which both sides of the issue endeavor to apply the powers of logic and persuasion support previously made emotional conclusions.
>
> 3. There is little empirical evidence which suggests that tests have certain consequences, either positive or negative....it is difficult if not impossible to single out one variable as the cause of other variables.
>
> 4. There are clear misuses of standardized achievement tests....when they are treated as an exact index of pupil achievement, when scores are used as the sole basis in making important decisions about individuals and programs, when users confuse the constructs that tests measure, and finally, when a single score is interpreted in terms of factors which caused that score.[20]

Finally, the Standards for Educational and Psychological Tests provides a framework for development of a testing program rationale and standards of operation. The Standards govern test selection and administration, reliability and validity, interpretation of test scores and

problems and issues in test use. Brown (1980) in his user's guide and commentary on the Standards expresses the hope that competent, effective and ethical use of tests requires qualified persons, using valid methods in an appropriate manner, to attain specified goals. As Brown states, "the goals of any assessment procedure should be to preserve the dignity of individuals and foster the development of their potentials." 21

Currently underway is the redesign and/or confirmation of the Rationale for the Testing Program and other elements of a compliance plan for submission to the Federal District Court. These include a Test Review Committee, the plan for monitoring development and use of Classroom Tests, the process for administration, scoring and interpretation of standardized test results, staff development program for use of standardized test results and development of individual student narrative reports. These will comprise elements of the Testing and Tracking component which will ensure non-discriminatory development, administration scoring and use of standardized and classroom tests in the Cleveland School District.

- NOTES -

1. REED vs. RHODES, C-73-1300, February 6, 1978, Remedial Order, 73.

2. Op. cit., 74.

3. Dyer, H. W. Recycling the problems in testing. Proceedings of the 1972 Invitational Conference on Testing Problems. Princeton, N. J.: Educational Testing Service, 1973.

4. Polemeni, A. Security in a city-wide testing program. Measurement in Education, 1975, 6, 3.

5. Mayo, S. T. Pre-service preparation of teachers in educational measurements. Chicago: Loyola University, 1967.

6. Bradley, R. The teaching of testing. Measurement News, 1978, 21, 4.

7. Clark, K. B. Prejudice and Your Child. Boston: Beacon Press, 1955.

8. Coleman, J. S., Campbell, E. Q., Hobson, C. J., McPartland, J., Mood, A. M., Weinfeld, F. D. & York, R. L. Equality of Educational Opportunity. Washington, D. C.: U. S. Government Printing Office, 1966.

9. Bradley, L. A. & Bradley, G. W. The academic achievement of black students in desegregated schools: A critical review. Review of Educational Research, 1977, 47, 399-499.

10. Hawley, W. D., Crain, R. L., Rossell, C. H., Fernandex, R. R., Schofield, J. W., Smylie, M. A., Tompkins, R., Trent, W. T. & Zlotnik, M. S. Strategies for effective desegregation: A synthesis of findings. Nashville: Center for Education and Human Development Policy, Institute for Public Policy Studies, Vanderbilt University, 1981, 115.

11. Mercer, J. R. & Lewis, J. F. The system of mylticultural pluralistic assessment (SOMPA). New York: The Psychological Corporation, 1979.

12. Elford, G., Fremco, J., Kates, R. & Rice, K. Handbook of management standards for awarding high school diplomas. Boston, Mass.: Massachusetts Advisory Council on Education, September 1975, VII.

13. Houts, P. L. (Ed.) The myth of measurability. New York, N. Y.: Hart, 1977.

14. Anderson, B. L. and others. Training citizen groups on educational testing issues: A trainer's manual. Washington, D. C.: National Institute of Education, September 1980.

15. Herndon, E. B. Your child and testing. Washington, D. C.: National Institute of Education, October 1980.

16. Gallagher, B. G. (ed.). NAACP report on minority testing. New York: National Association for Advancement of Colored People, May 17, 1976.

17. Flaugher, R. L. Bias in testing: A review and discussion. ERIC Clearinghouse on Tests, Measurement & Evaluation. Princeton, N. J.: Educational Testing Service, December 1974.

18. Merz, W. R. Methods of assessing bias and fairness in tests. Sacramento, Calif.: Applied Research Consultants, March 1980. (ERIC ED 198 155)

19. Berk, R. A. Handbook of methods for detecting item bias. Baltimore: The Johns Hopkins University Press, 1982.

20. Airasian, P. W. A perspective on the uses and misuses of standardized tests. Measurement in Education, 1979, 10, 13, 11-12.

21. Brown, F. G. Guidelines for test use: A commentary on the standards for educational and psychological tests. Ames, Ia.: National Council Measurement in Education, 1980, 114.

ROLE OF TEST REVIEW COMMITTEE

Marian Karantanes
Cleveland Public Schools

Paper presented to the annual meeting of the National Council on Measurement In Education, New York, New York, March, 1982.

ROLE OF TEST REVIEW COMMITTEE

The tests and testing procedures of the City-Wide Testing Program are based on recommendations developed by a Test Review Committee. Approximately every five years, such a committee has been convened to examine systematically all available and current standardized tests in order to formulate recommendations for the subsequent year's testing program. In determining the tests to be recommended, the committee considers the practical and technical features of each test, normative data, nondiscriminatory evidence, the correspondence with the curriculum and intended purpose, cost factors, etc.

Beginning with the 1979-1980 school year, the Test Review Committee was established as a standing committee in response to the Court Order regarding nondiscriminatory testing. Under the Court-approved standards for implementation of desegregation in the Cleveland Public School District, the School District shall have an internal assessment mechanism, procedure or system to identify problems and to initiate corrective action with regard to actual or possible discrimination or resegregation produced by testing or student placements. Such assessment mechanism, procedure or system shall be identifiable, and it shall have documented results.

The Test Review Committee currently is comprised of twenty-eight school, supervisory and administrative personnel. These staff are representative of the school levels and range of curriculum areas included in the City-Wide Testing Program. In accordance with the agreements maintained between the School District and the Cleveland Teachers' Union and the Cleveland Council of Administrators and Supervisors, each professional organization named committee members as their group representatives.

The Committee has met on an ongoing basis to review and monitor the City-Wide Testing Program and testing practices, recommend revisions in the testing program, and ensure maximum appropriate use of test results.

Key activities during the 1979-1980 school year included the review of: standardized instruments used in the City-Wide Testing Program to determine if any items reflect bias, use of test data from the City-Wide Testing Program to determine if there is misuse of results, and newly-developed standardized instruments in order to formulate recommendations for the subsequent year's testing program. Workshop sessions conducted by external consultants focused on identifying and selecting reliable, valid, nondiscriminatory tests and on multicultural, pluralistic assessment.

The study of bias was continued during the 1980-81 school year with the assistance of internal (school system) and external (nationally known) resource persons. The Committee's major accomplishments included:

- Recommending changes in Standardized Testing Administrative Regulations.
- Designing plan for review of tests--standardized and teacher-made.
- Exploring ways to identify and reduce bias in tests.
- Obtaining feedback from school staffs regarding key dimensions of the City-Wide Testing Program, such as:
  - Timeliness of delivery of test materials, test results
  - Usefulness of test results
  - Sequence and scope of testing program
  - Test security measures
- Expanding the information base regarding student characteristics and assessment procedure for placement of students in special educational programs such as: Learning Disabilities, Major Work, Educable Mentally Retarded, and LAU Compliance.

- Expanding the information base on student test-taking skills.
- Reviewing results of a City-Wide Testing Program Survey:
  - Elementary principals, teachers, and students
  - Secondary principals, guidance counselors, department chairpersons, teachers, students, and parents

The survey of school staff, parents and students was undertaken in April and May, 1981. Findings revealed differences between elementary and secondary practices but general concurrence among staff opinions about the system. Specifically:

- All elementary principals and a significant majority (93%) of secondary principals review standardized test results.
- All elementary principals and chairpersons/counselors provide for test security.
- A significant majority of elementary teachers (89%) and secondary teachers/counselors (85%) inform pupils of the purpose for taking standardized tests.
- A significant majority of elementary school staff (84%) and secondary school staff (85%) received inservice training and/or materials on interpretation/use of test results.
- Approximately 79% of the elementary school staff and 74% of the secondary school staff feel there is a relationship between the standardized test content and curriculum objectives.

Key outcomes of the student/parent survey include findings that:

- Approximately 82% of elementary students, 62% of secondary students and 36% of the parents were informed of the purpose for taking tests.
- Approximately 59% of elementary students, 27% of secondary students and 13% of the parents were provided with standardized test results.
- Approximately 90% of elementary students and 34% of secondary students expressed positive feelings about taking standardized tests.

Recommendations for improving the effectiveness of the Elementary Testing Program included:

- Intensify inservice program for school staff on test administration and interpretation of test results.
- Provide an inservice program for school staff on improving student test-taking skills.
- Provide an Individual Pupil Test Report for students and parents.
- Stagger tests in order to avoid heavy end-of-year test schedule.

Recommendations to improve the effectiveness of the Secondary Testing Program included:

- Academic supervisors should consider scheduling of standardized tests at earlier dates so that results can be returned to schools during semester tests are given.
- Test Review Committee function should continue to include evaluation of test content and curriculum objectives.
- Reporting of test results should include student attendance data.
- Alternate forms of tests should be purchased whenever funds are obtainable.
- Expansion of workshops that address the interpretation and use of test results should be continued.
- Designated test days should be scheduled with students arriving at specified times for standardized tests.
- Individual Pupil Test Reports should be provided for students and parents.

The Test Review Committee assumed responsibility for conducting the survey to obtain information regarding key dimensions of the City-Wide Testing Program. A total of five hundred ninety-two respondents representing one hundred twenty-two schools (seventy-eight elementary, twenty-eight junior high and sixteen senior high) completed the survey.

Committee efforts during the current 1981-1982 school year have focused on implementing on an ongoing basis a model for monitoring the City-Wide Testing Program to ensure nondiscriminatory test administration and use of test results in accordance with the Court Order. Key elements of this plan have included:

- Development of interview forms for principal, chairperson, counselor, teacher, pupil and parent--later revised to address committee members' problems/concerns which surfaced in the monitoring operation;
- Development of an observation instrument--which committee members deemed necessary as they implemented their monitoring tasks;
- Identification of fifteen test administration operations --seven at elementary, four at junior high, and four at senior high level;
- Identification of ten elementary schools per test, six junior highs per test, and four senior highs per test (using a random sample);
- Determination of specific committee member responsibilities.

Information gained from monitoring test-giving and test-taking activities in selected schools and from interviewing school staff members, students, and parents is then exchanged by committee members at monthly meetings. The monitoring plan appears to be effective in providing information/strategies to detect bias in tests and testing practices in the City-Wide Testing Program.

DESCRIPTION OF WORKSHOP USED WITH TEACHERS AND ADMINISTRATORS TO INTERPRET STANDARDIZED TEST RESULTS

Jacqueline Reed
Cleveland Public Schools

Paper presented to the annual meeting of the National Council on Measurement in Education, New York, New York, March, 1982.

DESCRIPTION OF WORKSHOP USED WITH TEACHERS AND ADMINISTRATORS TO INTERPRET STANDARDIZED TEST RESULTS

An extensive program of group administered survey type standardized tests is maintained by the Cleveland Public Schools, to provide a meaningful source of information about the adopted curriculum and overall student achievement. The testing program, selected by the Test Review Committee, embraces the cooperative efforts of administrators, curriculum and instructional supervisors, school principals, guidance counselors, teachers, the Department of Research, Development and Evaluation and the Division of Systems and Computer Services.

Under the Court-approved plans for a program of nondiscriminatory educational testing, the school system has as its goals to:

- Identify strengths and areas needing improvement in the current curriculum and instruction
- Compare achievement of students and student groups both within the district and with appropriate large city and national normative groups as means to evaluate student growth
- Provide diagnostic instructional information about individuals and groups of students to appropriate educational personnel
- Make available to parents and students information about students' performance on various test instruments
- Provide one basis for longitudinal study of achievement
- Provide an ongoing inservice training program for appropriate educational personnel

This paper addresses a segment of the inservice program provided to school staffs in the area of interpretation of standardized test results. Specifically, this presentation describes the rationale, preparation and implementation of a workshop designed to assist school personnel in their understanding and utilization of the computer printouts of Standardized Test Results.

Teachers and administrators from kindergarten to grade twelve receive computer printout data of standardized tests results. These printouts are locally generated with comparisons made to urban and national norms. The data include: alphabetical lists, class and school frequency reports, item response reports, skill-cluster reports, and stanine-banding reports with sex categories. In keeping with the desegregation order, racial/ethnic information was added to the stanine banding reports.

Prior to the court mandate, a conceptual framework for inservicing staffs in the computer printouts, based on school identified needs, was developed. Some basic assumptions about teachers and their professional development were contained in the framework. These assumptions included the view that teachers should:

- perceive the inservice as meaningful and purposeful
- receive feedback on their own behavior to make efficient use of experiences in learning
- participate in experiential learning experiences
- interact with other colleagues as a source for their own stimulation and growth
- evaluate their experiences to provide data for follow-up, additions, modifications and deletions in inservice

Peter and Schnare, (1976) studied the variables that affect inservice and concluded that, "a critical determinant of workshop success was the relevances of inservice content to subsequent classroom application".

Under the court approved plans, this inservice rationale continues to exist with refinements and expansions to meet the diversity of new and unforseen needs. The design of the workshop, approached from the logic that adults, like students, want hands-on and practical application of materials, utilizes such techniques as gaming, feedback, group interaction, discussion and evaluation. A review of the literature supports the notion that direct sensory involvement is an important element of inservice education. Harris, (1980) identifies activities such as "first hand" experiences, guided practice,

problem solving, analyzing and calculating as having a much greater experience impact on the learner than lecturing, panel discussions, films, television and filmstrip viewing. Bishop, (1976) indicates that research regarding teaching-learning styles, interaction analysis and instructional gain, suggests the importance of inquiry, gaming, stimulation, discovery, problem solving and "hands on" activities.

At the beginning of the school year a description of the inservice workshop on Use and Interpretation of Computer Printouts of Standardized Test Results is circulated to desegregation cluster directors, school principals, supervisors and administrators. Because of school staff willingness and school administration cooperation, all requests are handled. Inservice presentations are usually no more than an hour, because of budgetary and scheduling restraints. Within these limitations the following outline details the planning and procedures for this workshop.

PLANS AND PROCEDURES FOR WORKSHOP ON USE AND INTERPRETATION OF COMPUTER PRINTOUTS OF STANDARDIZED TEST RESULTS

I. PREPARATION

A. Resources and Materials

1. Call contact person. Set up date and time. Request participants list, overhead projector and screen.

2. Select a classroom sample of the alphabetical list, frequency report, item response report and skill cluster report.

   Attach to each printout questions that can be addressed quickly.

3. Prepare overhead transparencies for each sample printout and other selected material.

4. Develop a score sheet to record the number of correct responses of each team for each printout used.

5. Provide treats for entire group.

6. Design registration and evaluation forms to collect documentation for the court.

B. Format

1. Generate teams from participant's list blending grade level or subject area personnel.

2. Arrange room to provide for the number of teams generated. Display team names at each table.

3. Select places to distribute materials and check answer sheets.

II. IMPLEMENTATION

A. Registration

B. Introduction

1. Goal: Participants will be able to identify and use essential test information.

2. Objective: Participants will be able to identify and use data on the alphabetical list, frequency distribution report, item response report and item skill cluster report.

C. Procedures

1. Explain that each team is composed of people from various grade levels. Working together, they are to compete with the other teams in answering questions attached to each of the four basic printouts.

2. Direct each team to select two people, one who will act as the runner and the other who will act as a recorder for the team.

3. Demonstrate that the recorder will tear off the questions from the printout, fill in the team number and record the team answers. The runner will bring the team answer sheet to the facilitator and pick up the next printout. This process will continue until all printouts are addressed. Time limit for an hour's workshop--approximately 20 minutes.

4. Correct and record all team answer sheets. Return all answer sheets as soon as the team has completed its last printout or when time is called. Maintain a master score sheet to assess strengths and weaknesses of the group.

5. Present visualized lecturette of each printout emphasizing those areas where the groups showed weakness according to the score sheet.

6. Announce the winning team. Distribute treats to participants.

D. Evaluation

Data collected during the period of September 1980 to June 1981, of the total testing inservice program, including topics on use/interpretation of test data, construction/use of teacher-made tests and identifying and selecting reliable, valid, nondiscriminatory tests, revealed that 95 sessions were held with a total of 2,096 participants.

As determined by a sample of 305 rating sheets, 45% of the participants rated the "overall usefulness" of the inservice sessions as "excellent" and 41% as "above average". Some typical comments were: "Very clear, interesting presentation, more sessions like this needed"; "Enyoyable workshop"; "Well delivered informative session, but would be more welcome not at the end of a tired day"; "More inservice needed

on a regular basis".

While there is strong support for testing inservice sessions, scheduling and budgetary restraints continue to be the over-riding factors affecting implementation.

References

1. Berman, Barbara and Friederwitzer, Fredda. "A Pragmatic Approach to Inservice Education". Journal of the Association of Teacher Educators, Winter 1981, pp. 51-56.

2. Bishop, Leslee J. Staff Development and Instructional Improvement: Plans and Procedures. Boston: Allyn and Bacon, Inc., 1976.

3. Harris, Ben M. Improving Staff Performance Through In-Service Education. Boston: Allyn and Bacon, Inc., 1980

4. Peters, R. and Schnare, G. "Variables That Determine the Effectiveness of Teacher Inservice Workshops". Final Report to the National Institute of Education. Washington, D.C., May 1976

USE OF STANDARDIZED TEST RESULTS TO IMPROVE INSTRUCTION

Janie L. Rollins

Department of Research, Development and Evaluation

- Cleveland Public Schools -

Paper Presented at the Annual Meeting
of the American Educational Research Association
on March 21, 1982, New York, N.Y.

USE OF STANDARDIZED TEST RESULTS TO IMPROVE INSTRUCTION

The Cleveland Public Schools, Department of Research, Development and Evaluation, is fulfilling the mandate on Testing and Tracking of the Remedial Order of February 6, 1978, by providing appropriate inservice in interpretation of test scores to staff at all grade levels.

Results from the Cleveland city-wide testing program are intended to provide information to principal, counselors and teachers, so that they may be used in a nondiscriminatory manner to plan instruction more effectively. Being able to interpret the test scores properly and to use them in evaluation and instruction is high priority.

The task of making numerical data meaningful to the school staff is critical to the success of the entire testing program. Testing, a procedure for collecting relevant information, is valueless if used in isolation in the evaluation process. Observation, classroom performance, rating scales, interviews and testing are all part of this process. When data from the evaluation process is not translated so that it may be used by the teacher to improve instruction, then it is of little use to the teacher. Fleming (1979) stated that, "We must be able to translate measurement data into teaching prescription. The ultimate goal of a student evaluation program is to assist students in the development of their ability."

Information is needed to aid teachers in selecting various competing instructional approaches. This includes stated objectives arranged in hierarchical order for sequential learning and tests to determine students' patterns of responses. While results from achievement tests do not contain enough items for high reliability, response patterns of pupils scoring in stanines 1, 2, and 3 might be useful in revealing diagnostic

information. Mercer (1978) among others, relates that all tests should be used in a diagnostic manner; as an example, The CTBS Reading Test, and achievement test is given to Cleveland Public School students in the spring. Teachers are inserviced to use results of this test in combination with other data to provide instructional activities which will direct their students toward mastery of skills. Results must be used as a tool in charting students' progress and planning educational programs. With information from data of these test results, teachers can probe intensively into their students' needs. Test scores from standarized tests may represent information which can be used to compare students' progress with that of the norm group. The score must be recongized as one sample of performance at one point in time. They should assist in evaluating class achievement in terms of widely accepted educational objectives.

The testing division of the Cleveland Public Schools, Department of Research, Development and Evaluation provides inservice to school staff, teachers of the same grade or individual teachers. In this paper we shall discuss the method that is used in these inservice sessions, to assist teachers in translating test results data from the item response and item cluster records of a sixth grade CTBS Reading Test into prescriptions for teaching.

At grade level meetings, teachers are required to bring their class print-out sheets of alphabetical list, item responses and skill cluster results, school permanent record cards, and health cards.

The alphabetical printout is reviewed to identify students with scores in stanine 1, 2 or 3 in a particular skill. Longitudinal study of the students with a pattern of failures of test (grades K-present) in the same skills show indications that these skills have not been mastered. If

teachers verify that this information is a true picture of the student's performance, it is suggested that they work with reading specialist to design ways and means to help the student.

School records of these students are checked for normal or abnormal progress. Besides test scores, they contain other important information that might give the teachers insight into the students functioning or lack of functioning such as:

- remedial reading or math
- attendance regularity
- school changes
- teacher changes
- behavioral problems

The Item Response Record is provided for each class. Items are grouped according to content of a specific skill area in which they have been classified. The report list items for the comprehension test according to content within four specific areas, for example:

Skill	Content
Recognition	Articles Stories
Translation	Articles Stories Poems
Interpretation	Articles Stories Poems
Analysis	Stories Articles Poems

The Item Response Record reports the percentage of students in the class who answered the item correctly or ommitted it. As a class percentage

of response for a content area of a specific skill falls below 40 percent correct, it would suggest that attention should be given to this skill. Performance of 60 percent and above indicates strong response for the class. The Item Response Record also reports school and city percentages to aid teachers in comparing the level of correct response in the school and in the city. As class groups place 15 or more percent below the level of performance exhibited in their school and the city, it suggests that the class is performing below their peer groups.

The percentage of omits is furnished for each item for the class. When omit percentages are above 20 percent, the class generally needs practice in dealing with test items on a timed basis.

The Item Skill Cluster Record summarizes the item-by-item information in the Item Response Record. It gives the average percent of correct responses (for class, school, and city) for the cluster of items in each skill area.

The Item Response and Item Skill Cluster Records may be used by teachers to identify the strengths and weaknesses of their class in the skill areas. By examining the correct responses on these records teachers can determine skills that have been mastered by their class as well as those that need reinforcement.

If there has been a health problem (sight-hearing etc.) and it hasn't been corrected, it is suggested that teachers relate this information to the parents or school nurse.

The following design is a description of the method used during an inservice session:

INSERVICE DESIGN

ACTIVITY	PROCESS	FOLLOW-UP
Item Response Report A- Vocabulary	Circle words that 40% or less of the class had correct responses. . Discuss how words were used in the test item Compare words missed by the class with vocabulary taught at that grade level	Suggest working with reading specialist as how to build vocabulary base.
	Discuss percentage of class that omitted items . Check where omitted items are located on the test, beginning or end	Suggest proctoring class during testing. Notice how students schedule time allotted for test.
Item Response Report B	Identify the following: . number and name of skills . name of skills . content area . omits . percentage of correct response for class, school, and city	Ask teachers to consider relevance of skill to curriculum. Suggest teachers prescribe instructional activities for reinforcement.

INSERVICE DESIGN (Cont'd)

ACTIVITY	PROCESS	FOLLOW-UP
	Compare percentage of correct responses from the class with school and city.	
Item Skill Cluster Report	Identify and discuss the following: . skill categories . number of items in each skill . class, school, and city strongest skill . class, school, and city weakest skill . number of omits	IBID
Alphbetical Report	Identify students in stanine 1, 2 or 3. Check these students school record cards for previous test scores and other data	Suggest group according to needs.

Suggestions are made to form temporary groups to reteach skills that are not mastered. The alphbetical print-out can assist teachers in generating groups for this reinforcement. When skills have been mastered groups should be dissolved.

From February, 1981 to February 1982, 20 workshops for teachers have been presented by this writer.

The participants evaluations of the workshops are as follows:

. Number of staff served	170	
. Excellent	111	65%
. Above Average	55	32%
. Average	2	01%
. Poor	0	
. Not Applicable	2	01%
TOTAL	170	

COMMENTS FROM PARTICIPANTS

"More time -- Test results are two-fold -- teacher guidance and children's ability to follow directions."

"Enlightening"

"Very helpful with suggestions as to group (flexible) planning."

"I enjoyed the intimacy of our session, it allowed time for questions."

"Very helpful"

"Excellent presentation"

"Informational"

"Very enjoyable activity as well as interesting."

"Very interesting and helpful"

"Enjoyable, interesting as well as educational."

"Everything was made perfectly clear."

"This does not apply to me"

THE PUPIL NARRATIVE REPORT

Hazel D. Redden
Cleveland Public Schools

Paper presented to the annual meeting of the National Council on Measurement in Education, New York, New York, March, 1982.

THE PUPIL NARRATIVE REPORT

The Cleveland Public School System, under Desegregation Implementation since September, 1979, was directed by the Court to submit a compliance report to the Remedial Order that included planning and implementation results. The Testing Program of the Cleveland Public Schools is addressed in the July, 1981 Court-Shared Testing and Tracking Plan with the following goal statement:

Testing

> The Cleveland Public School System will maintain a program of nondiscriminatory individual and group testing that provides a source of information about the curriculum and overall student achievement.

Objectives with correspondent activities were established in order to process the implementation of the stated goal. Two objectives that are pending approval of the Court pertain to pupil narrative reports:

Objective

> Conduct a city-wide testing program as a means to analyze and demonstrate student achievement.

Objective

> Provide diagnostic instructional information about students' performance on various tests.

The following activities are designed to ensure that the rights of all students are properly protected.

Activities:

- Individual reports of test results will be distributed to teachers, parents, and students at the end of grades 3, 6 and 9 (entry to the next level of schooling), and will be available upon request at other grade levels.
- Information about the purpose of the testing program and how results are to be utilized will be provided to teachers, parents and students.

Activities (continued)

. Test data will be reported by school, cluster and school district.

The activity that involves the provision of individual test reports to teachers, parents and students has been studied for a number of years by test publishers as well as by the Cleveland Public Schools. It is only the excessive cost of individual test reporting that has prevented its inclusion in the city-wide testing program of Cleveland. An educational factor that is rarely tenable is judgment of one's performance. One of the aims in constructing an individual test report is to provide a design that would allow easy interpretation of scores for pupils and parents. Use of a narrative report form rather than numerical symbols in reporting test scores is more likely to initiate questions about reliability of the instrument used. As a consequence, messages that describe the performance level achieved on a test might be suspect. Nevertheless, the narrative report remains a favored format.

It would appear that reports of standardized test results to individual pupils and parents are now an integral part of the scoring services offered by test publishers[1]. Reports of this type are described as excellent aids in the identification of an individual's strengths and weaknesses as well as in planning appropriate instructional activities to meet the needs of pupils. The format used in the report that is sent to the home is one that invites parental involvement in aiding the student to meet educational objectives, a strong selling point for school systems. The provision of these reports affords the individual a form of

[1]Addison-Wesley Catalog of School Tests and Services, 1981, pp 27-30
CTB/McGraw-Hill Catalog: Test Materials and Scoring Services, 1982, pp 126-129
Psychological Corporation: 1982 Resources fo Decision Making - Catalog from the Test People, pp 26-27
Riverside Publishing Company Tests and Scoring Services Catalog, 1980-81, pp 13-16
Scott, Foresman and Company Test Division: Tests and Scoring Services, 1980-81, p 17

test reporting that utilizes a verbal format and eliminates almost all numbers and technical terms. Thus, a need identified over the years, the need for universal comprehension of test results, is met. Pupils, parents and teachers are better served with a style of interpretive feedback from standardized tests that is easily understood.

A paper, presented at the Mid-South Research Association meeting and sponsored by Iowa Testing Programs, Iowa City, Iowa, dealt with computer-generated narrative-format testing reports. The topic was "A Symposium on an Alternative to Traditional Test Reporting" (1972). The informational needs of pupils and parents were recognized then as needs for an alternative reporting system to serve the examinee and others who receive reports of test scores. The report encompassed related projects and was described in Resources in Education (1974).

Cleveland Public Schools, in acknowledgement of the need for systematic feedback of pupil performance on standardized tests, conducted a pilot program in 1978. On a pilot basis, an individual pupil test report was provided to a sixth grade class of 30 pupils with Comprehensive Tests of Basic Skills (CTBS) Math, Level 2 used as the instrument. Subtests were Computation, Concepts, and Applications (Appendix A). Eleven of the 30 parents attended an articulation meeting.

Findings revealed that the report provided adequate feedback of pupil performance on CTBS Math subtests. Resulting recommendations included the following:

(1) simplification of some terms to facilitate readability

(2) wider distribution of the test report

(3) simplified short note for parents to accompany test report

(4) reiteration that parent/teacher communication must occur.

A subsequent program was conducted at a Cleveland junior high school in 1981

as part of the continued plan to provide test results information to pupils and parents from the city-wide testing program. Participants were an English class consisting of 20 pupils. The instrument used was Comprehensive Tests of Basic Skills (CTBS) Reading, Level 4. Subtests were Vocabulary and Comprehension (Appendix B). Only two of the 20 parents attended the articulation meeting which was held on a Saturday morning. There was no exhaustive attempt to extend the experiment beyond the group that had been identified.

Findings from subsequent questionnaires revealed that there was positive reaction to the format of the report (Appendix C). The most common response was the comment that the report clarified the strengths and/or weaknesses of the pupils in specific content areas. In some instances, pupils thought their performance not commensurate to their ability. Where this occurred, pupils established test-taking improvement as their objective. Pupils and parents liked the individualized report and stated no difficulty with its readability. The general consensus was that an individual pupil report should be generated for every test administered in the city-wide testing program. One observation by a parent was that changes were needed to eliminate possible miscomprehension of stanine information. This concern will be considered at any future time of use of individual reports of test results. It should be noted, however, that during the articulation meetings held for parents, a stanine chart was distributed along with the pupil's test report (Appendix D). The stanine chart illustrates the normal curve and test performance categories and was reviewed with those parents in attendance.

The pilot programs conducted by the Cleveland Public Schools were exploratory and sensitive to the needs of pupils as documented from previous surveys. The

experiments indicate that if implementation of pupil narrative reports on a city-wide scale were to occur, the selection of the instrument would be based on the Rasch model. Since the pupil and parent are presented with an item by item response in its narrative format, complete with stanine description, the report form would be able to indicate the achievement of the pupil called a latent trait, i.e. a characteristic of the student which can be measured but is not directly observable. The second parameter of measurement, borrowed from the Rasch model, is reflected in the item description as well with the reporting of the performance of the pupil on each item. The second Rasch assumption is that the success a student has in answering a test question depends on his achievement level and the difficulty level of the item. This overview of the Rasch measurements can be found in the paper, "The Theory of the Rasch Model (For School Testing Folk)" by Fred Forster.

It is conceded that literature from major test publishers indicated the individual report of test results as critical to the complete service offered to the examinee. It is also conceded that with necessary funding, the City-Wide Testing Program of Cleveland Public Schools would include a pupil narrative report as part of the provision of basic testing services.

The Court Order addresses the issue of the Testing Program with the rationale that individual reports of test results should be incorporated into diagnostic instructional information. Reports should reflect student performance on standardized tests administered at the end of grades 3, 6 and 9, which are grades representive of entry to the next scholastic levels. The reports would be generated with narrative messages formulated to serve the unique needs of the examinee, the parents and teachers. Whereas, the pilot groups were small enough in size to

manually originate the verbal form and eliminate technical terms, a city-wide program would demand the use of computer services. The technology required would become cost effective once documentation and implementation of new tests occurred.

CTBS Comprehensive Tests of Basic Skills

(APPENDIX A)

INDIVIDUAL PUPIL TEST REPORT

To The Pupil:

This folder presents results from selected CTBS Tests which you took recently to find out which basic skills you are acquiring at satisfactory level and with which skills you need more help. The CTBS series was designed to measure skills related to the daily use of language and numbers.

This report results from testing during one particular week; there may have been reasons why it was not the best time for you to take the tests. If you feel that this report does not represent your idea of your work or if you have questions and want further explanation, you are encouraged to see your teacher or counselor.

DIVISION OF RESEARCH AND DEVELOPMENT

Cleveland Public Schools

ANALYSIS OF TEST ITEMS

MATHEMATICS COMPUTATION 48 ITEMS

	Item Numbers/Response Codes*										
Addition	1	2	3	4	5	6	7	8	9	10	11
	+	+	+	+	+	+	+	-	-	+	+
Subtraction	13	14	15	16	17	18	19	20	21	22	23
	+	+	+	+	+	+	+	+	+	+	-
Multiplication	25	26	27	28	29	30	31	32	33	34	35
	+	+	+	-	+	+	+	-	-	+	-
Division	37	38	39	40	41	42	43	44	45	46	47
	+	+	-	+	+	-	-	+	-	+	-

MATHEMATICS CONCEPTS 25 ITEMS

Recognize Curves, Shapes	10	13	14	15							
Compute Measurement	-	±	+	-							
Interpret Information on a Graph	3	4	5								
	+	+	-								
Convert Form	1	2	6	7	8	9	12	16	17	18	19
Compute Measurement	+	+	-	+	+	-	-	+	-	+	+
Understand Relationships	22	23	25								
	-	-	+								
Recognize Time	11	21	24								
Select Method to Solve Problem	+	+	-								

MATHEMATICS APPLICATIONS 25 ITEMS

Convert Form	26	27	28	29	30	31	32	33	34	35	36
Compute Measurement	-	-	+	-	-	-	+	+	+	-	+
Understand Set Relations	38	39	41	42	43	45	47	48	49		
Select Method to Solve Problem	-	-	-	-	+	+	-	+	-		
Determine Missing Terms	-										
Select Method to Solve Problem	40	44	46	50							
Determine Missing Terms	-	-	-	+							

---

* + CORRECT ITEM - INCORRECT ITEM O OMITTED ITEM

CTBS MATHEMATICS TESTS

LEVEL 2, FORM S

ANTONIO WYLEY ◀ PUPIL SCHOOL ▶ ANTHONY WAYNE

SIXTH ◀ GRADE TEACHER ▶ MRS. SARA CLARK

SEPTEMBER 26, 1977 ◀ DATE OF TESTING

MATHEMATICS TEST	Number of Items	Number Right	STANINES								
			Below Average			Average			Above Average		
			1	2	3	4	5	6	7	8	9
COMPUTATION	48	35						x			
CONCEPTS	25	15					x				
APPLICATIONS	25	10				x					

EXPLANATION OF CHART

TEST -- The name of each test which was taken.

NUMBER OF ITEMS -- This shows the number of items on the test which required an answer.

NUMBER RIGHT -- This tells the number of items answered correctly. It has no meaning by itself but is the key to the STANINE score.

STANINE -- Stanines are scores ranging from a low of 1 to a high of 9, with 5 representing typical performance. This score tells how you compare with other pupils in your grade.

The following general recommendations can be made on the basis of your test scores.

COMPUTATION: Your score shows slightly above average progress compared with other pupils in the same grade in the test norms. You answered most of the addition, subtraction, and multiplication problems correctly. Check with your teacher for additional practice with division problems.

CONCEPTS: Your score shows average progress compared with other pupils in the same grade in the test norms. Check with your teacher for ways to continue growth in this test area.

APPLICATIONS: Your score shows average progress compared with other pupils in the same grade in the test norms. Additional practice is needed in the area of problem solving. Check with your teacher.

DESCRIPTION
OF
COMPREHENSIVE TESTS OF BASIC SKILLS

READING

VOCABULARY TEST measures knowledge of meaning of words selected from various subject areas.

COMPREHENSION TEST measures understanding of reading material presented in sentences, paragraphs, letters, stories, articles and poems.

LANGUAGE

MECHANICS TEST measures skill in punctuation and capitalization.

EXPRESSION TEST measures correctness and effectiveness of expression.

MATHEMATICS

COMPUTATION TEST measures the four fundamental mathematical operations: addition, subtraction, multiplication and division.

CONCEPTS TEST measures the ability to change concepts from one form to another, and the understanding of numerical concepts.

APPLICATIONS TEST measures the ability to select and carry out problem-solving operations.

DESCRIPTION OF COMPREHENSIVE TESTS OF BASIC SKILLS

READING

VOCABULARY TEST measures knowledge of meaning of words selected from various subject areas.

COMPREHENSION TEST measures understanding of reading material presented in sentences, paragraphs, letters, stories, articles and poems.

LANGUAGE

MECHANICS TEST measures skill in punctuation and capitalization.

EXPRESSION TEST measures correctness and effectiveness of expression.

MATHEMATICS

COMPUTATION TEST measures the four fundamental mathematical operations: addition, subtraction, multiplication and division.

CONCEPTS TEST measures the ability to change concepts from one form to another, and the understanding of numerical concepts.

APPLICATIONS TEST measures the ability to select and carry out problem-solving operations.

CTBS Comprehensive Tests of Basic Skills

(APPENDIX B)

INDIVIDUAL PUPIL TEST REPORT

To The Pupil:

This folder presents results from selected CTBS Tests which you took recently to find out which basic skills are performed at a satisfactor level and which skills require more help. The CTBS series was designed to measure skills relate to the daily use of language and numbers.

This report results from testing during one particular week; there may have been reasons why it was not the best time for you to take the tes If you feel that this report does not represent your idea of your work or if you have questions and want further explanation, you are encouraged to see your teacher or counselor.

DEPARTMENT OF RESEARCH, DEVELOPMENT AND EVALUATION

Cleveland Public Schools

CTBS READING TESTS
LEVEL 4, FORM S

Any Student ◀ PUPIL SCHOOL ▶ Any School

9 ◀ GRADE TEACHER ▶ Any Teacher

3/16/81 ◀ DATE OF TESTING

Your score in the Vocabulary Test is in the (Below Average) (Average) (Above Average) range. Your score in the Reading Comprehension Test is in the (Below Average) (Average) (Above Average) range. See recommendations below.

READING TEST	Number of Items	Number Right	STANINES Below Average 23% 1	2	3	Average 54% 4	5	6	Above Average 23% 7	8	9
VOCABULARY	40	18					X				
COMPREHENSION	45	20					X				
Percent of Pupils in Norms			4%	7%	12%	17%	20%	17%	12%	7%	4%

EXPLANATION OF CHART

STANINE -- Stanines are scores ranging from a low of 1 to a high of 9. This score tell how you compare with other pupils in your grade. Stanines 1, 2 and 3 represent a below average band of performance compared with other pupils in the try-out (norms) group for the test. Twenty-three percent of the norms group falls in to this category. Stanines 4, 5 and 6 represent average performance. Fifty-four percent of the norms group falls into the average category. Stanines 7, 8 and 9 represent above average performance. Twenty-three percent of the norms group falls into the above average category.

RECOMMENDATIONS:

VOCABULARY -- Your score on the Vocabulary Test shows average progress compared with other students in the same grade on the test norms in knowing the meaning of words found in school reading materials. This means that your overall achievement among ninth graders is better than 40 percent of ninth graders nationally and that approximately 60 percent scored as well or better. Your teacher can suggest possible ways of continuing vocabulary growth.

COMPREHENSION -- Your score on the Comprehension Test shows average understanding of written materials presented in the test compared with other students in the same grade. This means that overall you scored better than 40 percent of ninth graders nationally and that approximately 60 percent scored as well or better. It is essential that you be encouraged to expand reading interests. Your school library has a variety of selections which will

ANALYSIS OF TEST ITEMS

READING VOCABULARY 40 ITEMS

Item Numbers/Response Codes*

Recall of Synonym - select word with same meaning as underlined word.	1	2	3	4	5	6	7	8	9	10	11	12	1
	-	+	+	+	-	-	+	-	+	-	-	+	
	15	16	17	18	19	20	21	22	23	24	25	26	2
	-	+	+	-	+	-	-	-	+	-	-	+	
	29	30	31	32	33	34	35	36	37	38	39	40	
	-	+	-	+	-	+	-	-	+	-	+	+	

Needs improvement in identification of word meanings

READING COMPREHENSION 45 ITEMS

Literal Recall answer who, what, where and when	17 24 38 - - +	Needs work in recall of details
Rewording answer questions by choosing reworded material	7 10 19 36 37 42 - + - + - +	Needs additional practice with reworded material
Context Clues determine meaning of word by its use	11 13 31 44 - - + -	Needs work in word meani
Main Idea Identify topic	1 14 22 25 28 35 39 41 45 - + - + - + - - +	Needs work in recognizing main topic
Descriptive Words describe persons, places or things	4 6 34 - - -	Shows weakness in descrip tion
Conclusions draw conclusions based on facts	2 5 8 9 12 15 16 18 20 29 40 - + - + + - + + + + -	Shows some strength in drawing of conclusions
Structural Style Identify writing and structural style	3 21 23 26 27 30 32 33 43 - + - - + + - + -	Shows some strength in recognizing writer's sty

* + CORRECT ITEM - INCORRECT ITEM O OMITTED ITE

(APPENDIX C)

INDIVIDUAL PUPIL TEST REPORT

CTBS, Level IV, Grade 9

1. Do you know what this report is? What do you think about it?

2. Did you look at your report?

_____Yes _____No

If yes, what kind of information did it give you about yourself that you did not know before?

3. What did your parents think about the report? What kind of comments did they make?

4. What part (parts) of this report was easy to understand?

5. What part (parts) of this report was hard to understand?

6. Your teacher explained to your class the content of this report. What did you understand from her explanation?

7. What part (parts) of her presentation was less clear?

PARENTS' RESPONSES (TELEPHONE AND SMALL GROUP DISCUSSIONS)

Individual Pupil Test Report

CTBS, Level IV, Gr. 9

1. What do you think about this report?

2. Miss S. Stewart, the English Teacher, gave a presentation about the contents of this report to the parents' group. What did you understand from her presentation? If you were not there, what did your child tell you about the report?

3. What part (parts) of her presentation was less clear to you?

4. Have you studied the report?

_____Yes _____No

If yes, what kind of information did this report reveal about your child that you did not know before?

5. What part (parts) of this report was easy to understand?

6. What part (parts) of this report was hard to understand?

7. How did you use this report to help your child?

8. Would you like to receive this type of report regularly?

_____Yes _____No

9. If this type of report was distributed to parents on a regular basis, do you think most parents would need a meeting in order to be able to understand the report?

10. Do you have suggestions to make this report easier to understand?

_____Yes _____No

If yes, How?

INDIVIDUAL PUPIL TEST REPORT

CTBS, Level IV, Grade 9

TEACHER'S QUESTIONNAIRE

Interviewer: _____ Date: _____

1a. What had you hoped to accomplish in providing this type of information to the pupils?

_____

_____

_____

_____

1b. Did you think that was accomplish in so far as this group of pupils is concerned?

_____

_____

_____

2a. What had you hoped to accomplish in providing the same type of information to parents?

_____

_____

_____

_____

2b. Do you think you accomplished your purpose with this group of parents?

_____Yes _____No

_____

_____

_____

3. How did the parents react to the presentation you made to them about this report?

_____

_____

_____

4. If you have to do it again, would you modify your presentation?

_____Yes _____No

If yes, how? _____

_____

5. What kind of comments/questions did you hear directly or indirectly about this type of report from the following :

a. School Staff: _____

_____

_____

b. Pupils (Gr. 9): _____

_____

_____

c. Parents (of pupils): _____

_____

_____

6a. What do you think of the report format? Which part (parts) communicated clearly?

_____

_____

_____

6b. Which part (parts) of the report did not appear to get the idea across? How would you improve it?

_____

_____

_____

INDIVIDUAL PUPIL TEST REPORT

CTBS, Level IV, Grade 9

PRINCIPAL'S QUESTIONNAIRE

Interviewer: _____ Date: _____

1a. What had you hoped to accomplish in providing this type of information to the pupils?

_____

_____

_____

_____

1b. Did you think that was accomplished in so far as this group of pupils is concerned?

_____

_____

_____

2a. What had you hoped to accomplish in providing the same type of information to parents?

_____

_____

_____

_____

2b. Do you think you accomplished your purpose with this group of parents?

_____Yes _____No

_____

_____

_____

3. How did the parents react to the presentation you made to them about this report?

_____

_____

_____

4. If you have to do it again, would you modify your presentation?

_____Yes _____No

If yes, how? _____

_____

5. What kind of comments/questions did you hear directly or indirectly about this type of report from the following :

a. School Staff: _____

_____

_____

b. Pupils (Gr. 9): _____

_____

_____

c. Parents (of pupils): _____

_____

_____

6a. What do you think of the report format? Which part (parts) communicated clearly?

_____

_____

_____

6b. Which part (parts) of the report did not appear to get the idea across? How would you improve it?

_____

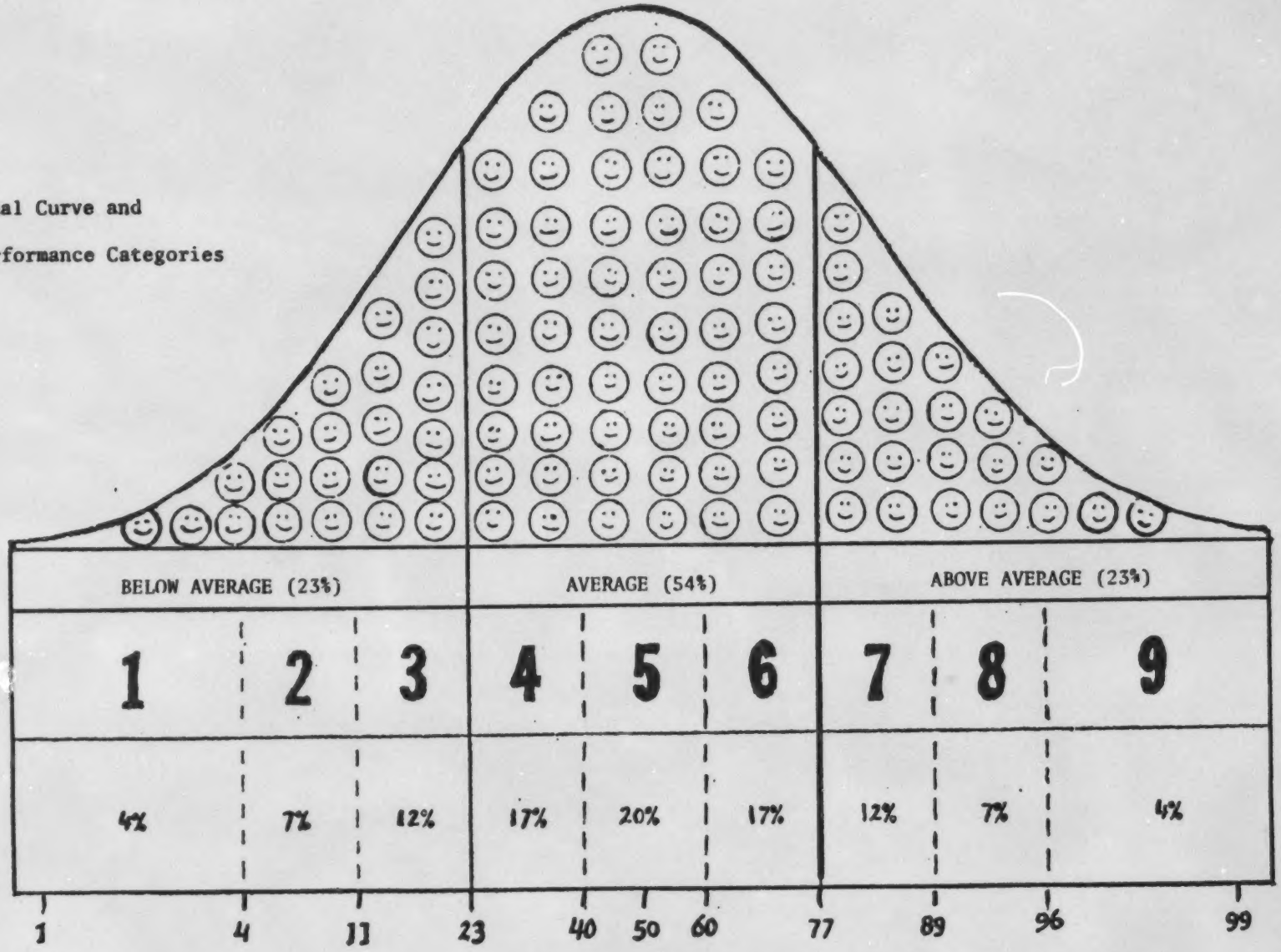
_____

_____

The Normal Curve and

Test Performance Categories

(APPENDIX D)



POSSIBLE DIAGNOSTIC PROCEDURES

STEPS	INSTRUMENT/TECHNIQUE
. Identification of children requiring teaching reinforcement	CTBS Reading Comprehension Test
. Development of Information Base about Children	Permanent Record Reading Record Health Record Teacher/Principal observations Classroom performance Attitude toward reading
. Planning reinforcement instruction for children	School resources?

Possible questions of concern:

1. Did children attend kindergarten?
2. What were children's levels of readiness from readiness tests, teacher observations?
3. What was progress in grade one, two, three? Steps completed?
4. Have children received any special services in reading? What were outcomes? Reading marks? Test scores?
5. Are there health problems? Sight and Hearing Status?
6. Do children attend school regularly?
7. What is latest Cognitive Abilities score? What are performance levels on various sub-test areas?
8. What have present teachers observed? Children's speech patterns, listening levels, attention span, special strengths, etc.
9. How well do children read orally? Silently on classroom assignments--books, chalkboard assignments? What is apparent level of mastery of phonics, sight words, etc.?

TESTING UNDER COURT ORDERED DESEGREGATION

Improving Teachers' Test Making Skills

Barbara Chambers

Cleveland Public Schools

Paper presented to the annual meeting of
the National Council on Measurement in
Education in Education, New York, New York,
March, 1982.

TESTING UNDER COURT ORDERED DESEGREGATION

Improving Teachers' Test-Making Skills

"...therefore, defendents are cautioned that they must insure that all tests, whether standardized, criterion-referenced, or teacher-made, are developed, administered, and scored in a nondiscriminatory manner." (Remedial order, p. 74)

The major focus of attention on testing--whether from the media, educators or commerical test publishers--generally, has been over the years on standardized tests. Questions regarding fairness, difficulty, length, bias, administration, reliability, validity of tests always are directed toward standardized tests. The direction of the court order to insure that teacher-made tests are developed, administered and scored in a nondiscriminatory manner, introduced a new dimension of responsibility to the administrator of our school district. Traditionally, teacher-made tests have been considered the bastion of teachers' academic freedom; many teachers have been reluctant to share tests with other teachers! The task of complying with this component of the Remedial Order truly would be a challenging, yet an exciting one!

To provide direction to the school district in complying with all aspects of the Remedial Order, the court issued mandated guidelines. These guidelines addressing teacher-made tests were:

- hold workshops for teachers on constructing and using teacher-made tests
- develop a procedure to maintain continual quality-control checks on teacher-made tests

The plans developed by the Department of Research and Development, guided by the framework of these two items regarding teacher-made tests, will be presented here.

To place the development of the teacher-made testing compliance plans in perspective, it is useful to review some background information. Over the years, workshops in constructing teacher-made tests had been available from the Department of Research and Development. The workshops were usually designed to meet the specific needs of particular departments or groups of teachers, such as curriculum developers, federal project test developers, or the like. Although school administrators recognized teachers' need for assistance in constructing teacher-made tests, the opportunity for district-wide provision of such inservice training was limited by several factors. Among them were the lack of financial resources and the lack of available useful teacher-training materials and strategies.

Helping to alleviate these difficiencies was a grant from ESEA, Title IV-C. In 1976, the Department of Research and Development received funding to develop a Teacher-Training Package to improve teachers' knowledge and skills in classroom testing. A sixteen-member team of teachers assisted in the developmental process by reacting to materials and strategies designed by the project manager.

Resulting from a two-year development effort was the CLASSROOM TEST SKILLS Teacher-Training Package. The Package consists of three slide/tape programs to be used as advance organizers for four units of teacher-centered activities. The four units are:

- Planning The Test
- Constructing Test Items
- Assembling and Administering The Test
- Scoring and Analyzing The Test

The package is designed to be presented by an instructional leader to teachers who use paper and pencil test in four two-hour inservice sessions. It was validated by the State of Ohio Department of Education.

This teacher-training package developed by the Classroom Test Skills project provided resources for the school district in developing a plan to address the court mandate to provide workshops in constructing and using nondiscriminatory teacher-made tests. The plan, implemented in the spring of 1980, included the following elements:

- All teachers received a 40-page booklet titled Resource Materials Excepted from a CLASSROOM TEST SKILLS.
- All teachers were directed to view the three CLASSROOM TEST SKILLS slide/tape programs on educational T.V. (Video tapes were made available for additioned in-school presentations.) An average of 3,697 (83.5%) viewed each program.
- Workshops in nondiscriminatory classroom testing were made available to all schools and departments on a continuous basis.

All aspects of this plan were monitored and regularily reported to the court.

Guidelines for preparing classroom tests included in the Classroom Test Skills materials gave impetus to the development of a test-rating instrument. This rating instrument formed the key element in the plan developed to comply with the second court mandate regarding teacher-made test, the maintenance of a quality-control check.

This plan, now completing its second year of implementation, includes the following procedures:

1. Each semester every teacher submits one major (chapter or units) teacher-made test to a file maintained by the principal in elementary schools or department chairpersons in secondary schools.
2. Principals (elementary) or department chairpersons (secondary) submit the most recent grade level/subject tests file to subject supervisors upon request. The collected test files are then forwarded to the Department of Research, Development and Evaluation. (Test files from six grade levels or specific subjects are generally reviewed during each school year.)

3. The Department of Research, Development and Evaluation selects a random sample of tests from the submitted tests files. At least one test is selected from each school. Tests are not identified by teacher or school, but are given a code number.

4. A committee reviews each test in the sample by using the rating instrument developed from the Classroom Test Skills guidelines. Membership on this committee generally includes the supervisor, two persons from the Department of Research, Development and Evaluation, and two or more classroom teachers or department chairpersons.

5. The test ratings are compiled and a report is prepared by the Department of Research, Development and Evaluation. This summary report, the rating sheet and the test file are returned to the individual schools.

6. Principals and department chairpersons are directed to review the rated tests and summary report with teachers with the purpose of improving teacher-made test construction. Additional inservice workshops maybe requested from Research, Development and Evaluation.

Since this process of monitoring the quality-control of teacher-made tests was initiated in the spring of 1980, the following subject areas have been reviewed:

- Upper Elementary Mathematics
- Senior High English
- Primary Language Arts
- 10th grade French
- Junior High Social Studies
- Senior High Mathematics
- Junior High Industrial Arts
- Senior High Science
- Elementary Social Studies
- Junior High English

Review of the Summary Reports resulting from this process show the following key findings:

- Tests items included language and attitudes which were universal and non-discriminary.
- The majority of tests were easy to read and neat in appearance.
- Generally items were grouped by item-type and maintained the same format within the item-type.
- More than half of the tests did not include concise, complete direction.
- Point values for items were very seldom included on tests.
- Essay items are used very infrequently on teacher-made tests.
- Mathematics tests, as expected use primarily problem-type items; other subjects tend to rely on short-answer responses or multiple-choice items.
- The various item-types were generally constructed according to guidelines.

In the area of teacher-made tests, compliance with the Remedial Order has given the Cleveland School District the opportunity to implement a rather innovative process of analysis. While much work remains to be done in improving the quality of teacher-made tests, nationally, as well as in Cleveland, this focus of attention on teachers' tests has opened a new window into the classroom.